Human Genome Project

Dr. Aninda Mandal
Assistant Professor of Botany
A. B. N. Seal College
Cooch Behar 736101, WB

Introduction to HGP

- The HGP was an international scientific research project that aimed to determine the complete sequence of nucleotide base pairs that make up human DNA and all the genes it contains.
- It remains the world's largest collaborative biological project.
- The idea was picked up in 1984 by the US government when the planning started, the project was formally launched in 1990 and was declared complete in 2003.
- The human genome is by far the most complex and largest genome.
- Its size spans a length of about 6 feet of DNA, containing more than 30,000 genes.
- The DNA material is organized into a haploid chromosomal set of 22 (autosome) and one sex chromosome (X or Y).

Salient Features of Human Genome

- Human genome consists the information of 24 chromosomes (22 autosome + X chromosome + one Y chromosome); in *Homo sapiens* ($2n = 2x = 46$).
- The human genome contains **over 3 billion nucleotide pairs**.
- Human genome is estimated to have **about 30,000 genes**.
- Average gene consists of 3000 bases. But sizes of genes vary greatly, with the largest known human gene encoding dystrophin containing 2.5 million base pairs.
- Only about **3 % of the genome encodes amino acid sequences of polypeptides** and **rest of it junk** (repetitive DNA).
- The functions are unknown for over 50% of the discovered genes.

Salient Features of Human Genome

- The repetitive sequences makeup very large portion of human genome. Repetitive sequences have no direct coding function but they shed light on the chromosome structure, dynamics and evolution.
- **Chromosome 1 has most genes (2968)** and **Y chromosome has the lowest (231)**.
- Almost all nucleotide bases are exactly the same in all people. Genome sequences of different individuals differ for **less than 0.2% of base pairs**. Most of these differences occur in the form of single base differences in the sequence. These single base differences are called **single nucleotide polymorphisms (SNPs)**. One SNP occurs at every ~ 1,000 bp of human genome. About 85% of all differences in human DNAs are due to SNPs.

Goals of Human Genome Project

- To identify and map all the genes in human DNA.
- To develop a genetic linkage map of human genome.
- To obtain a physical map of human genome.
- To develop technology for the management of human genome information.
- To know the function of genes.
- Determine the sequences of the 3 billion chemical base pairs that make up human DNA.
- Store this information in public databases.
- Develop tools for data analysis.
- Allow the private sector access to the informations and technologies that arise from this project.

Participating countries and funding agencies

- In 1990, the 2 major funding agencies, the **US Department of Energy (DOE)** and **National Institute of Health (NIH)**, developed an MoU in order to coordinate plans and set the clock for the initiation of the Project. The **$3-billion** project was formally launched by the 2 agencies.
- Most of the government-sponsored sequencing was performed in **20 Universities and research centers** in the United States, the United Kingdom, Japan, France, Germany, Canada, and China.
- A parallel project was conducted outside the government sponsorship by the **Celera Corporation or the Celera Genomics** which was formally launched in 1998.

Pioneers in HGP

- **Robert Sinsheimer** proposed the idea of sequencing the human genome in the year **1985**.
- **Charles DeLisi and David Smith** proposed the **budget** for Human Genome Project.
- **HGP act** was passed in the **US congress under President Regan** in **1988**.
- **James Watson headed** the NIH Genome Program.
- **Francis Collins** succeeded James Watson in **1993** as the overall Project Head and the Director of the NIH (which later become the National Human Genome Research Institute NHGRI) and was in power until the completion of HGP in 2003.
- **Jim Kent**, a PhD scholar in the University of California Santa Cruz (UCSC), in May 2000, developed a software, **GigAssembler**, that allowed the publicly funded Human Genome Project to assemble and publish the human genome sequence

Timeline of HGP

- **1986**: The **birth** of the Human Genome Project.
- **1990**: Project **initiated** as joint effort of US Department of Energy and the National Institute of Health.
- **1994**: **Genetic Privacy Act**: to regulate collection, analysis, storage and use of DNA samples and genetic information is proposed.
- **1996**: **Welcome Trust joins** the project.
- **1998**: **Celera Genomics** (a private company founded by Craig Venter) formed to sequence much of the human genome in 3 years.
- **1999**: **Completion of the sequence** of **Chromosome 22 - the first human chromosome to be sequenced**.
- **2000**: **Completion** of the working draft of the **entire human genome**.
- **2001**: Analysis of the working draft are **published**.
- **2003** HGP sequencing is **completed** and Project is declared finished two years ahead of schedule.

The DNA sequence of human chromosome 22

I. Dunham, N. Shimizu, B. A. Roe, S. Chissoe *et al.*†

† *A full list of authors appears at the end of this paper*

Knowledge of the complete genomic DNA sequence of an organism allows a systematic approach to defining its genetic components. The genomic sequence provides access to the complete structures of all genes, including those without known function, their control elements, and, by inference, the proteins they encode, as well as all other biologically important sequences. Furthermore, the sequence is a rich and permanent source of information for the design of further biological studies of the organism and for the study of evolution through cross-species sequence comparison. The power of this approach has been amply demonstrated by the determination of the sequences of a number of microbial and model organisms. The next step is to obtain the complete sequence of the entire human genome. Here we report the sequence of the euchromatic part of human chromosome 22. The sequence obtained consists of 12 contiguous segments spanning 33.4 megabases, contains at least 545 genes and 134 pseudogenes, and provides the first view of the complex chromosomal landscapes that will be found in the rest of the genome.

Nature 1999;402:489-95

Contributors: I. Dunham[1], A. R. Hunt[1], J. E. Collins[1], R. Bruskiewich[1], D. M. Beare[1], M. Clamp[1], L. J. Smink[1], R. Ainscough[1], J. P. Almeida[1], A. Babbage[1], C. Bagguley[1], J. Bailey[1], K. Barlow[1], K. N. Bates[1], O. Beasley[1], C. P. Bird[1], S. Blakey[1], A. M. Bridgeman[1,*], D. Buck[1,*], J. Burgess[1,*], W. D. Burrill[1], J. Burton[1], C. Carder[1], N. P. Carter[1], Y. Chen[1], G. Clark[1], S. M. Clegg[1], V. Cobley[1], C. G. Cole[1], R. E. Collier[1], R. E. Connor[1,*], D. Conroy[1,*], N. Corby[1], G. J. Coville[1], A. V. Cox[1], J. Davis[1,*], E. Dawson[1], P. D. Dhami[1], C. Dockree[1], S. J. Dodsworth[1,*], R. M. Durbin[1], A. Ellington[1], K. L. Evans[1], J. M. Fey[1], K. Fleming[1], L. French[1], A. A. Garner[1], J. G. R. Gilbert[1], M. E. Goward, D. Grafham[1], M. N. Griffiths[1], C. Hall[1,*], R. Hall[1], G. Hall-Tamlyn[1], R. W. Heathcott[1,*], S. Ho[1,*], S. Holmes[1], S. E. Hunt[1], M. C. Jones[1], J. Kershaw[1], A. Kimberley[1], A. King[1], G. K. Laird[1], C. F. Langford[1], M. A. Leversha[1], C. Lloyd[1], D. M. Lloyd[1], I. D. Martyn[1], M. Mashreghi-Mohammadi[1], L. Matthews[1], O. T. McCann[1], J. McClay[1], S. McLaren[1], A. A. McMurray[1], S. A. Milne[1], B. J. Mortimore[1], C. N. Odell[1], R. Pavitt[1], A. V. Pearce[1], D. Pearson[1], B. J. Phillimore[1], S. H. Phillips[1], R. W. Plumb[1], H. Ramsay[1], Y. Ramsey[1], L. Rogers[1], M. T. Ross[1], C. E. Scott[1], H. K. Sehra[1], C. D. Skuce[1], S. Smalley[1], M. L. Smith[1], C. Soderlund[1], L. Spragon[1], C. A. Steward[1], J. E. Sulston[1], R. M. Swann[1], M. Vaudin[1,*], M. Wall[1], J. M. Wallis[1], M. N. Whiteley[1,*], D. Willey[1], L. Williams[1], S. Williams[1], H. Williamson[1,*], T. E. Wilmer[1], L. Wilming[1], C. L. Wright[1], T. Hubbard[1], D. R. Bentley[1], S. Beck[1], J. Rogers[1], N.Shimizu[2], S. Minoshima[2], K. Kawasaki[2], T. Sasaki[2], S. Asakawa[2], J. Kudoh[2], A. Shintani[2], K. Shibuya[2], Y. Yoshizaki[2], N. Aoki[2], S. Mitsuyama[2], B. A. Roe[3], F. Chen[3], L. Chu[3], J. Crabtree[3], S. Deschamps[3], A. Do[3], T. Do[3], A. Dorman[3], F. Fang[3], Y. Fu[3], P. Hu[3], A. Hua[3], S. Kenton[3], H. Lai[3], H. I. Lao[3], J. Lewis[3], S. Lewis[3], S.-P. Lin[3], P. Loh[3], E. Malaj[3], T. Nguyen[3], H. Pan[3], S. Phan[3], S. Qi[3], Y. Qian[3], L. Ray[3], Q. Ren[3], S. Shaull[3], D. Sloan[3], L. Song[3], Q. Wang[3], Y. Wang[3], Z. Wang[3], J. White[3], D. Willingham[3], H. Wu[3], Z. Yao[3], M. Zhan[3], G. Zhang[3], S. Chissoe[4], J. Murray[4], N. Miller[4], P. Minx[4], R. Fulton[4], D. Johnson[4], G. Bemis[4], D. Bentley[4], H. Bradshaw[4], S. Bourne[4], M. Cordes[4], Z. Du[4], L. Fulton[4], D. Goela[4], T. Graves[4], J. Hawkins[4], K. Hinds[4], K. Kemp[4], P. Latreille[4], D. Layman[4], P. Ozersky[4], T. Rohlfing[4], P. Scheet[4], C. Walker[4], A. Wamsley[4], P. Wohldmann[4], K. Pepin[4], J. Nelson[4], I. Korf[4], J. A. Bedell[4], L. Hillier[4], E. Mardis[4], R. Waterston[4], R. Wilson[4], B. S. Emanuel[5], T. Shaikh[5], H. Kurahashi[5], S. Saitta[5], M. L. Budarf[6], H. E. McDermid[6], A. Johnson[6], A. C. C. Wong[6], B. E. Morrow[7], L. Edelmann[7], U. J. Kim[8], H. Shizuya[8], M. I. Simon[8], J. P. Dumanski[9], M. Peyrard[9], D. Kedra[9], E. Seroussi[9], I. Fransson[9], I. Tapia[9], C. E. Bruder[9], K. P. O'Brien[9].

Addresses: 1, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 2, Department of Molecular Biology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan; 3, Department of Chemistry and Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 311, Norman, Oklahoma 73019, USA; 4, Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Blvd, St. Louis, Missouri 63108, USA 5, Division of Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia and the Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA; 6, Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada; 7, Department of Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA; 8, California Institute of Technology, Division of Biology, Pasadena, California 91125, USA; 9, Department of Molecular Medicine, Clinical Genetics Unit, Karolinska Hospital, CMM bldg. L8:00, 17176 Stockholm, Sweden

** Present addresses: Division of Virology, Department of Pathology, Tennis Court Road, Cambridge CB12 1QP, UK (A. M. Bridgeman); Lark Technologies, Radwinter Road, Saffron Walden Essex, CB11 3HY, UK (D. Buck, R. E. Connor, S. Ho); Australian Genome Research Facility, Gehrmann Laboratories, University of Queensland, St Lucia QLD 4072, Australia (J. Burgess, J. Davis); Incyte Europe Ltd, 214 Cambridge Science Park, Cambridge CB4 0WA, UK (D. Conroy); Tepnel Life Sciences, Innovation Centre, Scotscroft Building, Wilmslow Road, Didsbury, Manchester M20 8RY, UK (S. J. Dodsworth); Department of Brassica & Oilseeds Research, Cambridge Laboratory, John Innes Centre, Norwich, Norfolk, UK (C. Hall); Cancer Genetics Lab, Department of Biochemistry, University of Otago, P.O. Box 56, Dunedin, New Zealand (R. W. Heathcott); Dept. of Zoology, University of Cambridge, Downing Street, CB2 3EJ, UK (M. N. Whiteley); Monsanto, Genomic Team, Mail Zone N3SA, 800 North Lindbergh Boulevard, St. Louis, Missouri 63167, USA (M. Vaudin); Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK (H. Williamson)*

The DNA sequence of human chromosome 21

The chromosome 21 mapping and sequencing consortium
M. Hattori*, A. Fujiyama*, T. D. Taylor*, H. Watanabe*, T. Yada*, H.-S. Park*, A. Toyoda*, K. Ishii*, Y. Totoki*, D.-K. Choi*, E. Soeda†, M. Ohki‡, T. Takagi§, Y. Sakaki*§; S. Taudien‖, K. Blechschmidt‖, A. Polley‖, U. Menzel‖, J. Delabar¶, K. Kumpf‖, R. Lehmann‖, D. Patterson#, K. Reichwald‖, A. Rump‖, M. Schillhabel‖, A. Schudy‖, W. Zimmermann‖, A. Rosenthal‖; J. Kudoh☆, K. Shibuya☆, K. Kawasaki☆, S. Asakawa☆, A. Shintani☆, T. Sasaki☆, K. Nagamine☆, S. Mitsuyama☆, S. E. Antonarakis, S. Minoshima☆, N. Shimizu☆, G. Nordsiek††, K. Hornischer††, P. Brandt††, M. Scharfe††, O. Schön††, A. Desario‡‡, J. Reichelt††, G. Kauer††, H. Blöcker††; J. Ramser§§, A. Beck§§, S. Klages§§, S. Hennig§§, L. Riesselmann§§, E. Dagand§§, T. Haaf§§, S. Wehrmeyer§§, K. Borzym§§, K. Gardiner#, D. Nizetic‖, F. Francis§§, H. Lehrach§§, R. Reinhardt§§ & M.-L. Yaspo§§**

Consortium Institutions:
** RIKEN, Genomic Sciences Center, Sagamihara 228-8555, Japan*
‖ Institut für Molekulare Biotechnologie, Genomanalyse, D-07745 Jena, Germany
☆ Department of Molecular Biology, Keio University School of Medicine, Tokyo 160-8582, Japan
†† GBF (German Research Centre for Biotechnology), Genome Analysis, D-38124 Braunschweig, Germany
§§ Max-Planck-Institut für Molekulare Genetik, D-14195 Berlin-Dahlem, Germany
Collaborating Institutions:
† RIKEN, Life Science Tsukuba Research Center, Tsukuba 305-0074, Japan
‡ Cancer Genomics Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan
§ Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan
¶ UMR 8602 CNRS, UFR Necker Enfants-Malades, Paris 75730, France
 # Eleanor Roosevelt Institute, Denver, Colorado 80206, USA
*** Medical Genetics Division, University of Geneva Medical School, Geneva 1211, Switzerland*
‡‡ CNRS UPR 1142, Institut de Biologie, Montpellier, 34060, France
‖ School of Pharmacy, University of London, London WC1N 1AX, UK

Chromosome 21 is the smallest human autosome. An extra copy of chromosome 21 causes Down syndrome, the most frequent genetic cause of significant mental retardation, which affects up to 1 in 700 live births. Several anonymous loci for monogenic disorders and predispositions for common complex disorders have also been mapped to this chromosome, and loss of heterozygosity has been observed in regions associated with solid tumours. Here we report the sequence and gene catalogue of the long arm of chromosome 21. We have sequenced 33,546,361 base pairs (bp) of DNA with very high accuracy, the largest contig being 25,491,867 bp. Only three small clone gaps and seven sequencing gaps remain, comprising about 100 kilobases. Thus, we achieved 99.7% coverage of 21q. We also sequenced 281,116 bp from the short arm. The structural features identified include duplications that are probably involved in chromosomal abnormalities and repeat structures in the telomeric and pericentromeric regions. Analysis of the chromosome revealed 127 known genes, 98 predicted genes and 59 pseudogenes.

BBC | Home | News | Sport | Radio | TV | Weather | Languages | Search

Low graphics | Accessibility help

BBC NEWS

Watch One-Minute World News

News services
Your news when you want it

News Front Page

Africa
Americas
Asia-Pacific
Europe
Middle East
South Asia
UK
Business
Health
Science & Environment
Technology
Entertainment
Also in the news

Video and Audio

Programmes
Have Your Say
In Pictures
Country Profiles
Special Reports

RELATED BBC SITES
SPORT
WEATHER

Last Updated: Monday, 14 April 2003, 16:01 GMT 17:01 UK

E-mail this to a friend Printable version

Human genome finally complete

By Ivan Noble
BBC News Online science staff

The biological code crackers sequencing the human genome have said they have finished the job - two years ahead of schedule.

Their announcement came less than three years after a "rough draft" was published to worldwide acclaim.

When UK Prime Minister Tony Blair and then US President Bill Clinton hailed the publication of the draft in June 2000, 97% of the "book of life" had been read.

Decoding using the power of robotics and computers (*Image by The Wellcome Trust Sanger Institute*)

The decoding is now close to 100% complete. The remaining tiny gaps are considered too costly to fill and those in charge of turning genomic data into medical and scientific progress have plenty to be getting on with.

The Wellcome Trust Sanger Institute, the only British institution taking part in the international effort, completed almost a third of the sequence - the biggest contribution by a single institution.

VIDEO AND AUDIO NEWS

The BBC's Sue Nelson
"British scientists contributed almost one third of the human genome"

Watch

SEE ALSO

- 'Secret of life' discovery turns 50
  27 Feb 03 | Science/Nature
- DNA databases 'no use to terrorists'
  15 Jan 03 | Science/Nature
- Mouse clues to human genetics
  04 Dec 02 | Science/Nature
- Genome man to create new life
  21 Nov 02 | Science/Nature
- Scientists crack human code
  26 Jun 00 | Science/Nature

RELATED INTERNET LINKS

- DNA 1953-2003
- Nature: 50 years of DNA
- Wellcome Trust Sanger Institute
- NIH National Human Genome Research Institute
- Celera Genomics

The BBC is not responsible for the content of external internet sites

TOP SCIENCE & ENVIRONMENT STORIES

Importance

- With the HGP molecular medicine can improve diagnosis of disease, detect genetic predispositions to disease earlier, ration drug design, control gene therapy and control systems of drugs, and create pharamacogenomics customs drugs. The potential for using genes themselves to treat disease, known as gene therapy, has captured the imaginations of the public and the biomedical community for good reason. This rapidly developing field holds great potential for treating or even curing genetic and acquired diseases.
- Teach us how humans evolved and continue to evolve.
- Help identify criminals.
- Allow new biofuels to be made: Genomics may be a key factor in addressing the global energy crisis. Through their knowledge in this field, researchers are developing a better understanding of how to harness various renewable energy sources, such as lignocellulosic biomass, microalgae, and cyanobacteria. Furthermore, it appears that genetic engineering of enzymes will be a key factor in optimizing development of sustainable biofuels that can someday replace fossil fuels on a global scale.

Thank You